P1: Image Generation using Stable Diffusion and ComfyUI.

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning
with
TechSaksham – A joint CSR initiative of Microsoft & SAP

by

Zobiya Hussain,
hussainzobiya27@gmail.com

Under the Guidance of

Adharsh P & Jay Rathod

ACKNOWLEDGEMENT

I would like to take this opportunity to express my deep gratitude to all individuals who have supported me directly or indirectly during this project.

Firstly, I extend my heartfelt thanks to our supervisors, **Mr. Adharsh P & Mr. Jay Rathod**, for their invaluable guidance, insightful feedback, and continuous encouragement throughout this project. Their expertise and mentorship have played a crucial role in the successful completion of my work on **P1: Image Generation using Stable Diffusion and ComfyUI**.

I am also grateful to **TechSaksham, Microsoft, and SAP** for providing me with this opportunity and a platform to explore the advancements in AI and deep learning. Their support in terms of resources and technical assistance has been instrumental in our learning and research journey.

Additionally, I acknowledge the contributions of the research community whose work in the field of generative AI and diffusion models has inspired and guided us. I also appreciate our peers and colleagues for their collaboration and discussions, which have enriched my understanding of the subject.

Lastly, I extend my sincere gratitude to my family and friends for their unwavering support and encouragement, which have been a source of motivation throughout this endeavour.

ABSTRACT

This project explores **image generation using Stable Diffusion and ComfyUI**, two powerful AI-based tools for generating high-quality images from textual prompts. Stable Diffusion is a deep-learning model capable of generating photorealistic images through diffusion processes. ComfyUI, a modular and user-friendly interface, facilitates experimentation with various Stable Diffusion parameters and models.

The primary objective of this project is to implement a system that efficiently generates diverse and high-resolution images from textual descriptions. The methodology includes setting up Stable Diffusion, integrating it with ComfyUI, and optimizing parameters to enhance image quality.

The results demonstrate the effectiveness of Stable Diffusion in generating visually appealing images with different styles and themes. This project also discusses potential applications in creative design, content generation, and AI-assisted artistry.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1Problem Statement:

Image generation has seen significant advancements with deep learning, but generating high-quality and customizable images remains a challenge. This project explores the capabilities of **Stable Diffusion** with **ComfyUI**, allowing users to create highly detailed images efficiently.

1.2Motivation:

With the rise of generative AI, industries such as gaming, advertising, and content creation require **realistic AI-generated visuals**. This project investigates how **Stable Diffusion** can address this need by providing an interactive and user-friendly approach through **ComfyUI**.

1.3Objective:

- Implement Stable Diffusion for high-quality image generation.
- Utilize ComfyUI for a modular and user-friendly experience.
- Optimize diffusion parameters for improved image outputs.
- Evaluate generated images based on quality and diversity.

1.4Scope of the Project:

This project focuses on **text-to-image** generation, **image enhancement**, and **style transfer** using AI. Limitations include hardware constraints and the need for fine-tuning model parameters.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

The field of **generative AI** has evolved significantly with models like GANs, VAEs, and more recently, **diffusion models**. **Denoising Diffusion Probabilistic Models (DDPMs)**, introduced by Ho et al. (2020), laid the foundation for Stable Diffusion. Further developments in **Latent Diffusion Models (LDMs)**, proposed by Rombach et al. (2022), enabled more efficient and scalable implementations.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

- **GANs (Generative Adversarial Networks)**: Used in models like StyleGAN but suffer from mode collapse and training instability.
- **VAEs (Variational Autoencoders)**: Efficient but struggle with generating fine-grained details.
- **Diffusion Models**: Overcome many GAN and VAE limitations by progressively refining images from noise.
- **Stable Diffusion**: A latent diffusion model that balances efficiency and quality, supporting **text-to-image** synthesis.
- **ComfyUI**: Provides a **modular UI** that enhances experimentation with Stable Diffusion parameters.

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

- Computationally Expensive: Generating high-resolution images still requires powerful GPUs.
- Lack of Control: Some generative models lack fine-tuning capabilities for specific styles.
- Bias and Ethical Concerns: AI-generated images can reflect biases in training data.

How This Project Addresses These Gaps

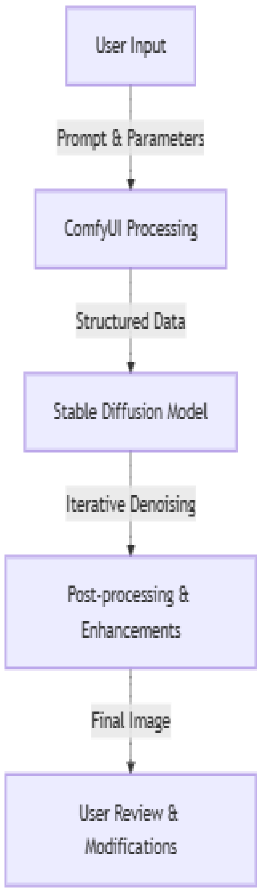
- Uses ComfyUI to provide greater control over image generation parameters.
- Optimizes diffusion steps and model settings to enhance performance on consumer hardware.
- Implements custom fine-tuning methods to allow for style and subject-specific image generation.

CHAPTER 3

Proposed Methodology

3.1 System Design

A workflow diagram showcasing the interaction between **Stable Diffusion, ComfyUI, and user inputs**.



3.2 Requirement Specification

3.2.1 Hardware Requirements:

- GPU (NVIDIA RTX 3060 or higher recommended)
- 16GB RAM or more
- High-performance CPU (Intel i5 or AMD Ryzen 5 and above)

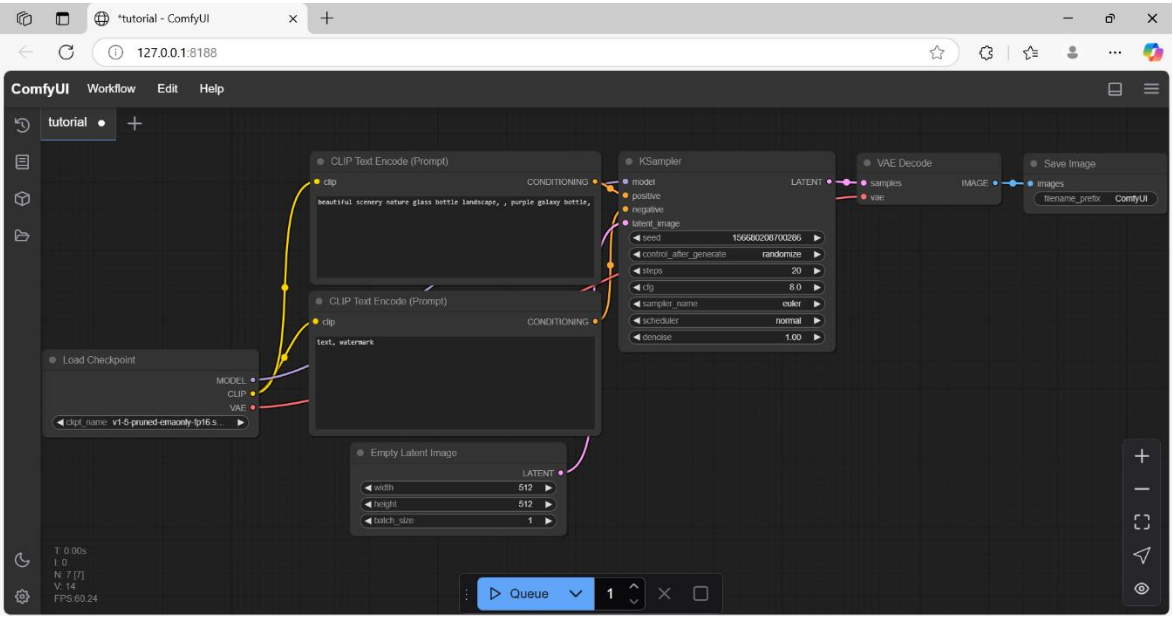
3.2.2 Software Requirements:

- Python 3.8+
- PyTorch, Hugging Face Diffusers, Transformers
- ComfyUI framework

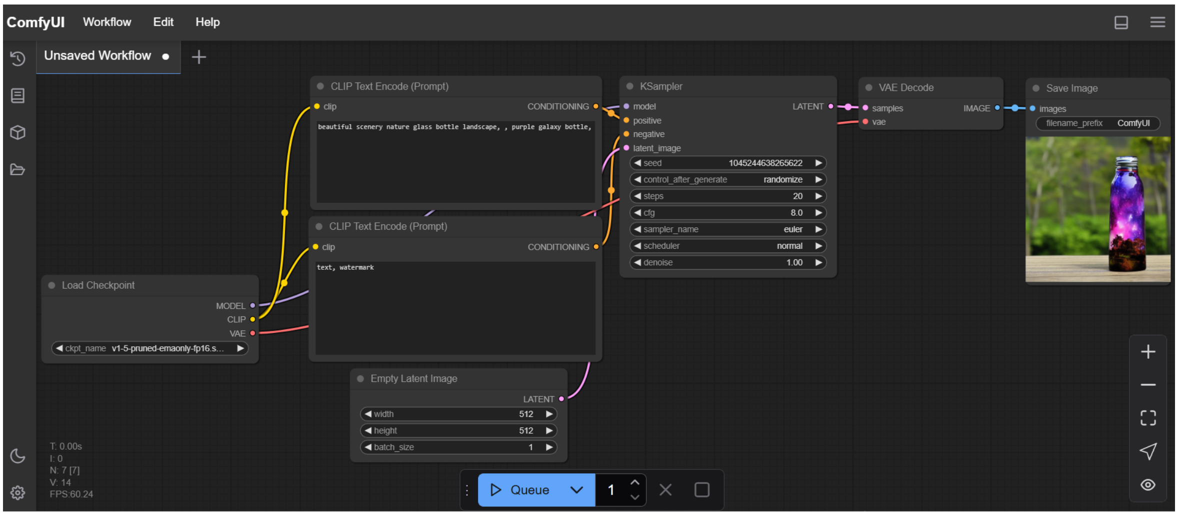
CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:



The above snapshot depicts the GUI of Comfy UI.



This snapshot shows the image generated on giving the prompt beautiful scenery nature glass bottle landscape, purple glass bottle.

The image generated is above:

4.2GitHub Link for Code:

https://github.com/ZobiyaH/StableDiffusion_ComfyUI_Project.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

- Enhancing image resolution using **super-resolution techniques**.
- Implementing **real-time generation** with optimized inference models.

5.2 Conclusion:

This project successfully demonstrates **Stable Diffusion with ComfyUI** for AI-based image generation.

REFERENCES

[1]. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, "Detecting Faces in Images: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.